

# You Only Look Once (YOLO)

2022.02.19.SAT

Minju Kim

# What is Object Detection?

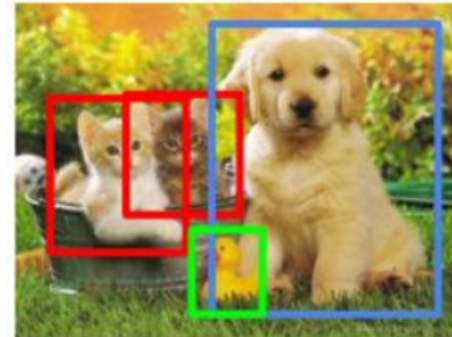
- **Object Detection** locates the presence of objects with a bounding box and types or classes of the located objects in an image
  - Input : An image with one or more objects, such as a photograph
  - Output : One or more bounding boxes(e.g. defined by a point, width, and height), and a class label for each bounding box.

**Classification**



CAT

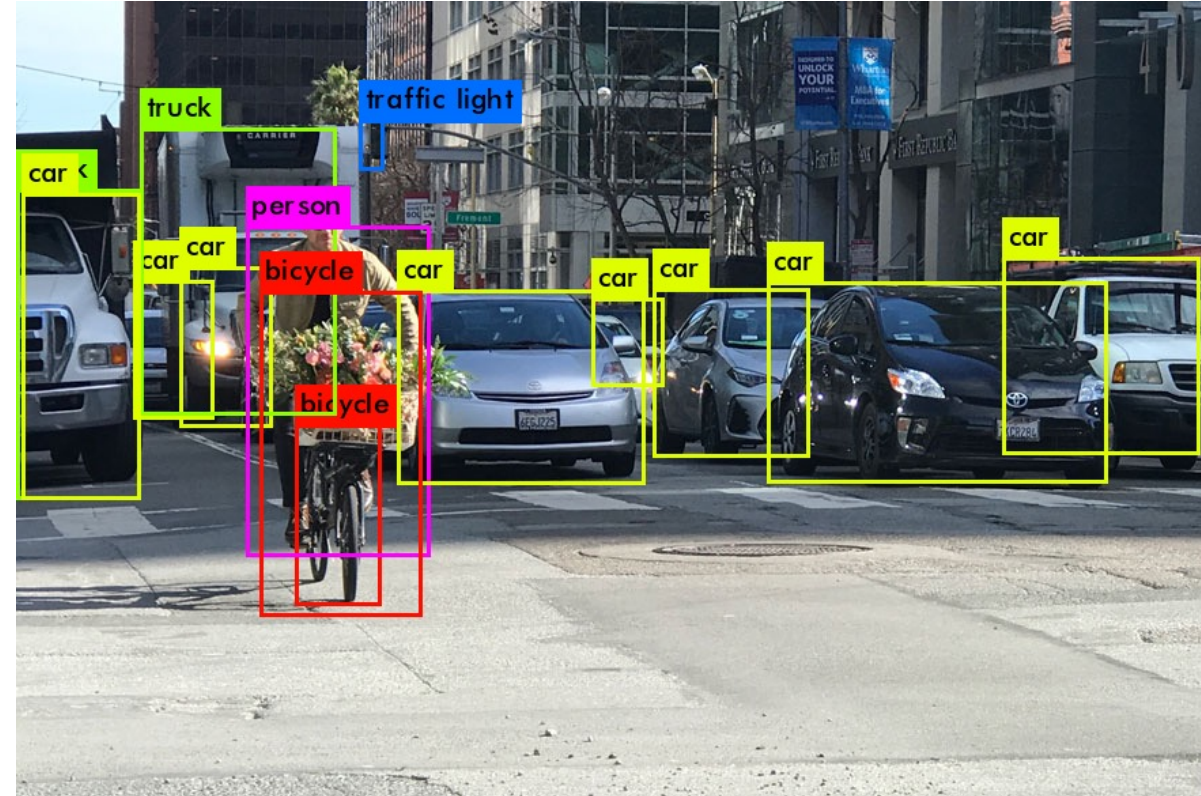
**Object Detection**



CAT, DOG, DUCK

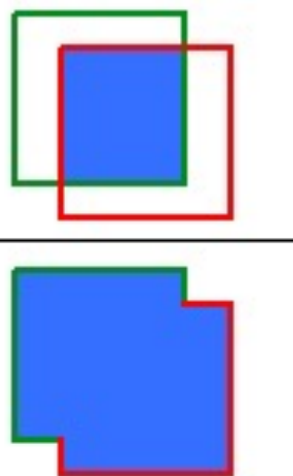
# What is You Only Look Once (YOLO) ?

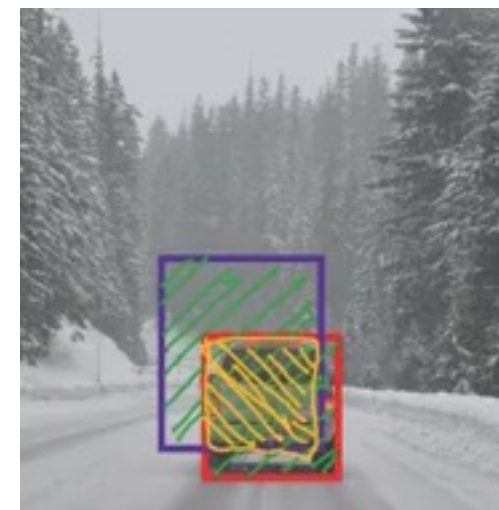
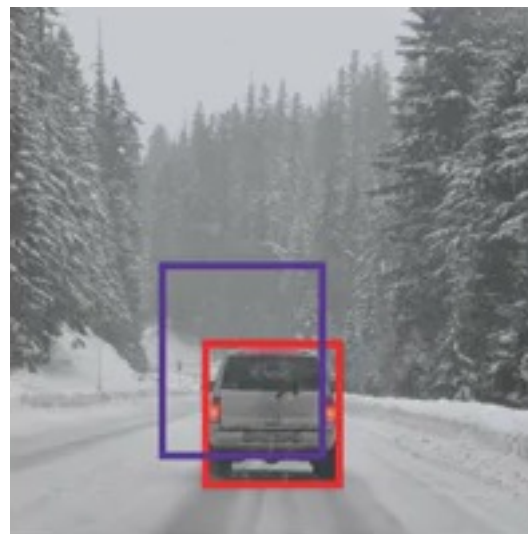
- **YOLO(You Only Look Once)**, is a network for Object detection.
- YOLO algorithm “only looks once” at the image in the sense that it only requires the image or the video to pass through the neural network once to make predictions.
- With YOLO, a single CNN simultaneously predicts multiple bounding boxes and class probabilities for those boxes.



# Terminology

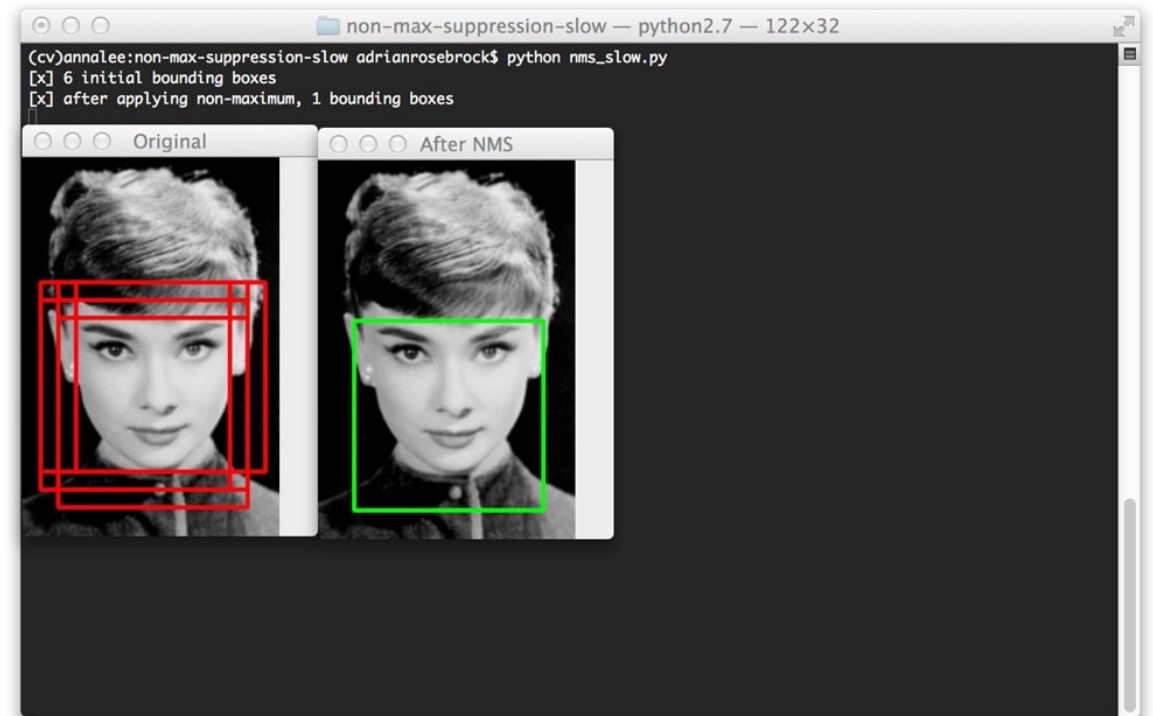
- Intersection over union (**IoU**)
  - The intersection divided by the Union
  - Gives us the ratio of the overlap to the total area, providing a good estimate of how close the bounding box is to the original prediction.
  - The IoU is how well the machine's predicting bounding box matches up with the actual object's bounding box.

$$IoU = \frac{\text{area of overlap}}{\text{area of union}} =$$




# Terminology

- NMS (Non-maximum suppression)
  - NMS ensures we identify the optimal cell among all of the bounding boxes
  - NMS chooses the highest probability of the boxes that are determining the same object





# How YOLO works?

Split the image into  $S \times S$  grids.



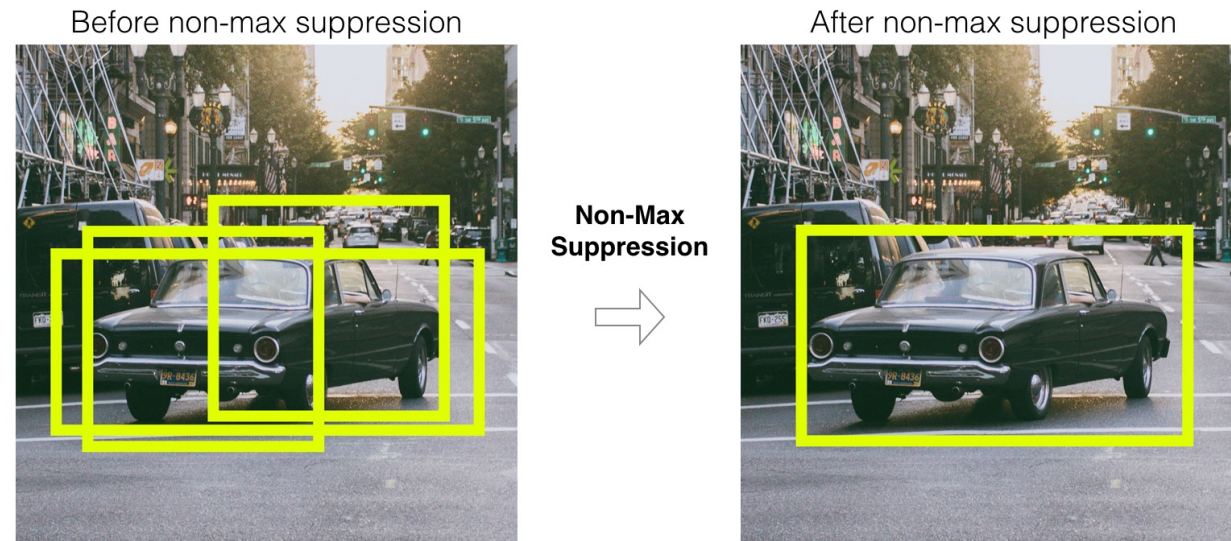
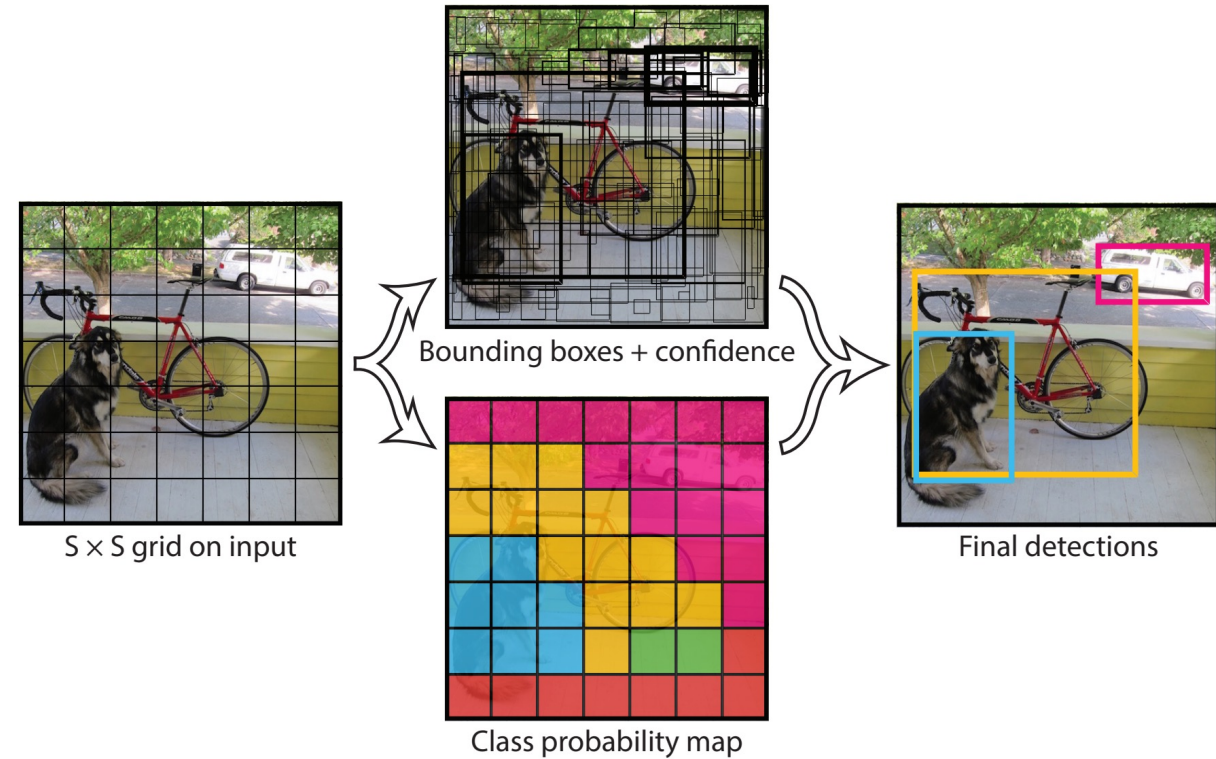
These grids predict bounding box coordinates relative to their cell coordinates, along with the object label and probability of the object being present in the cell.



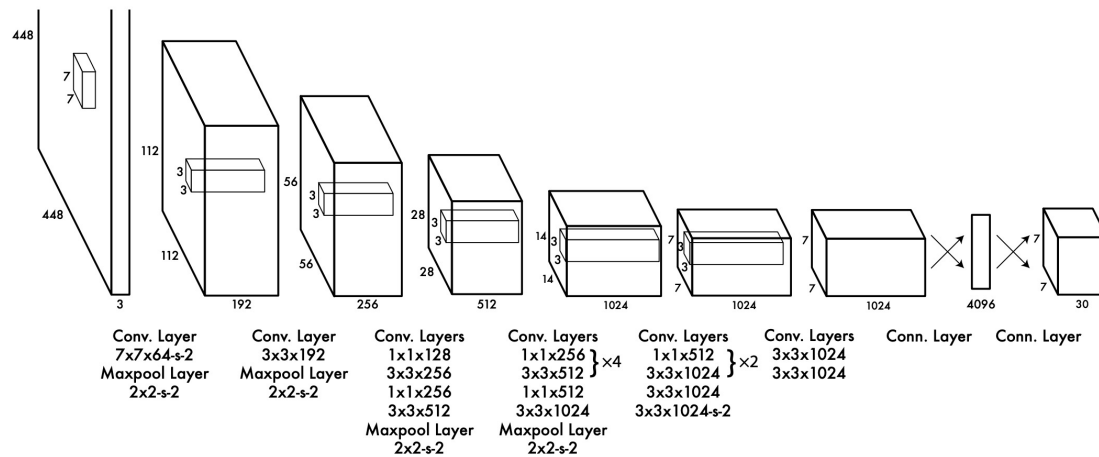
A lot of duplicate predictions due to multiple cells predicting the same object with different bounding box predictions.



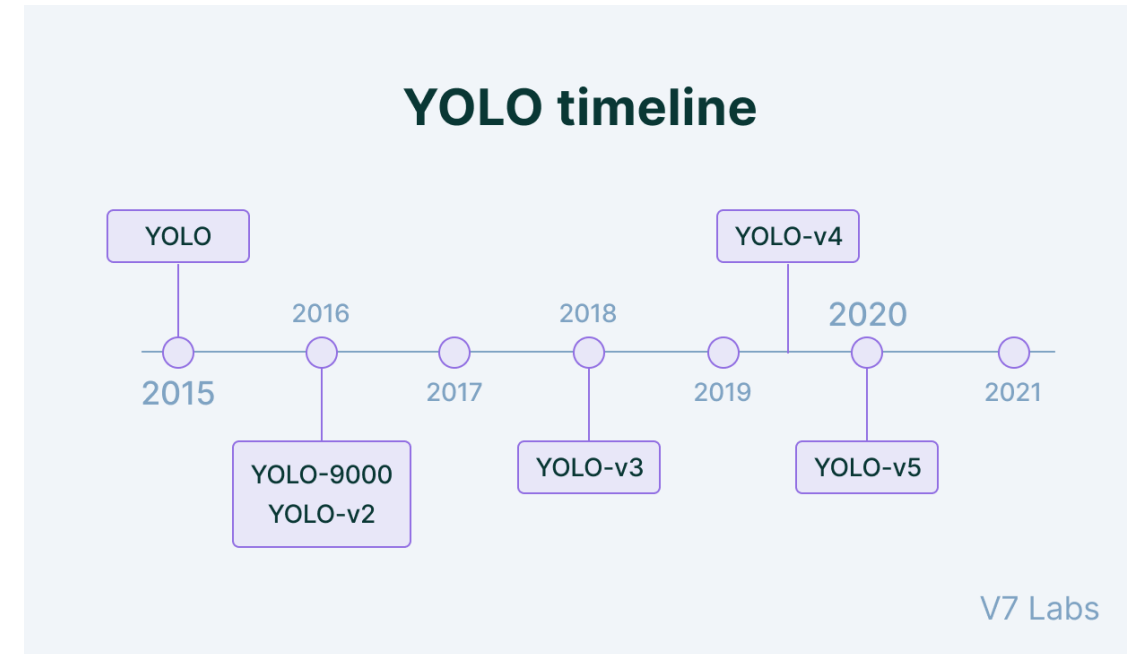
YOLO makes use of Non Maximal Suppression to suppresses all bounding boxes that have lower probability scores.



# YOLO Architecture & development



Inspired by the GoogleNet architecture, YOLO's architecture has a total of 24 convolutional layers with 2 fully connected layers at the end.



# Summary

- YOLO provided a super fast and accurate object detection algorithm
- Over 5 versions (3 official) and cited more than 16 thousand times
- Autonomous driving, vehicle detection, and intelligent video analytics.





# References

- <https://arxiv.org/pdf/1506.02640.pdf>
- <https://machinelearningmastery.com/object-recognition-with-deep-learning/>
- <https://hoyao12.github.io/blog/Tutorials-of-Object-Detection-Using-Deep-Learning-what-is-object-detection/>
- <https://medium.datadriveninvestor.com/overview-of-yolo-673f5019dof2>
- <https://www.v7labs.com/blog/object-detection-guide>
- <https://appsilon.com/object-detection-yolo-algorithm/>
- <https://medium.datadriveninvestor.com/overview-of-yolo-673f5019dof2>
- <https://www.youtube.com/watch?v=8DjIJc7xH5U&t=85s>
- <https://www.v7labs.com/blog/yolo-object-detection>

Thank you!